

## C L A I M S

1. A document searching method comprising:  
employing a computer to receive, from a user, a query including at least  
5 one search term;  
employing computerized answer retrieving functionality to generate  
document search terms including at least one additional search term not present in said  
query, which said at least one additional search term was acquired, prior to receipt by  
said computer of said query from said user, by said computerized answer retrieving  
10 functionality in response to at least one query in the form of a question; and  
operating computerized search engine functionality to access a set of  
documents in response to said query, based not only on at least one search term supplied  
by said user in said query, but also on said at least one additional search term provided  
by said computerized answer retrieving functionality.  
15
2. A document searching method according to claim 1 and wherein said  
query is a question.
3. A document searching method according to claim 1 and wherein said  
20 query is not a question.
4. A document searching method according to any of the preceding claims  
and wherein said employing computerized answer retrieving functionality provides said  
at least one additional search term by retrieving search terms, acquired other than in  
25 response to earlier questions, received by said computerized answer retrieving  
functionality prior to receipt of said query from said user.
5. A document searching method according to any of the preceding claims  
and wherein said employing a computer comprises employing said computer to receive  
30 said query by at least one of:  
typing said query;  
using a voice responsive input device;

using a screen scraping functionality;  
using an email functionality;  
using an SMS functionality; and  
using an instant messaging functionality.

5

6. A document searching method according to any of the preceding claims and wherein said employing computerized answer retrieving functionality to generate document search terms comprises utilizing computerized query normalizing functionality for normalizing said query.

10

7. A document searching method according to claim 6 and wherein said normalizing said query is performed based at least in part on at least one of a plurality of query normalization rules.

15

8. A document searching method according to any of the preceding claims and wherein said employing computerized answer retrieving functionality to generate document search terms comprises generating document search terms, including said at least one additional search term not present in said query by replacing at least one word in said query by at least one selected synonym thereof.

20

9. A document searching method according to claim 8 and wherein said replacing at least one word in said query by at least one selected synonym thereof comprises employing computerized synonym retrieving functionality to identify said at least one selected synonym at least partially by reference to at least one word in said query other than said at least one word which is replaced by said at least one selected synonym.

25

10. A document searching method according to claim 9 and wherein said employing computerized synonym retrieving functionality comprises identifying said at least one selected synonym by:

30

identifying a plurality of synonyms; and

selecting at least one of said plurality of synonyms for which there exists a phrase in a corpus which is relevant to said query.

11. A document searching method according to claim 10 and wherein said  
5 identifying said at least one selected synonym comprises:

searching said corpus for occurrences of at least one of said plurality of synonyms for which there exists a phrase in said corpus which is relevant to said query;  
and

designating at least one of said plurality of synonyms as a selected  
10 synonym in accordance with a number of occurrences in said corpus of a phrase including said at least one of said plurality of synonyms which is relevant to said query.

12. A document searching method according to any of the preceding claims and also comprising utilizing computerized query processing functionality to process  
15 said query prior to said operating computerized search engine functionality, said utilizing computerized query processing functionality including:

utilizing said computerized query processing functionality to generate at least one expected answer to said query;

utilizing said computerized query processing functionality to generate at  
20 least one preliminary search engine query based on said at least one expected answer;

utilizing said computerized query processing functionality to concatenate said at least one preliminary search engine query with said at least one additional search term not present in said query, thereby to form a concatenated search engine query; and

providing said concatenated search engine query to said computerized  
25 search engine functionality.

13. A document searching method according to any of the preceding claims and also comprising providing a representation of at least one document in said set of documents to said user.

30

14. A document searching method according to claim 13 and wherein said providing a representation comprises presenting at least one link to said at least one document.

5 15. A document searching method according to any of the preceding claims and also comprising:

extracting at least one answer to said query from at least one document in said set of documents; and

providing said at least one answer to said user.

10

16. A document searching method according to claim 15 and wherein said extracting at least one answer comprises analyzing said at least one document by:

carrying out theme extraction on said at least one document, said theme extraction utilizing statistical analysis of frequency of occurrence of words to identify at

15 least one theme word of said at least one document;

extracting sentences from said at least one document;

selecting at least one of said sentences as a potential answer;

scoring each of said at least one of said sentences selected as a potential answer; and

20 identifying said at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

17. A document searching method according to either of claims 15 and 16 and wherein said extracting at least one answer comprises:

25 enhancing said at least one document by:

identifying capitalized phrases which appear in said at least one document;

identifying designated capitalized words belonging to said capitalized phrases; and

30 adding, to said at least one document, adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, said

designated capitalized word that does appear alongside thereof elsewhere in said document in a capitalized phrase; and

carrying out analysis of said at least one document in order to identify at least one portion thereof as a potential answer.

5

18. A document searching method according to any of claims 15 to 17 and wherein said providing said at least one answer to said user comprises presenting said at least one answer in an editable report precursor format.

10 19. A document searching method according to any of the preceding claims and wherein said employing computerized answer retrieving functionality comprises employing artificial intelligence.

20. A system for document searching comprising:  
15 a computer operative to receive, from a user, a query including at least one search term;

computerized answer retrieving functionality operative to generate document search terms including at least one additional search term not present in said query, which said at least one additional search term was acquired, prior to receipt by  
20 said computer of said query from said user, by said computerized answer retrieving functionality in response to at least one query in the form of a question; and

computerized search engine functionality operative to access a set of documents in response to said query, based not only on said at least one search term but also on said at least one additional search term provided by said computerized answer  
25 retrieving functionality.

21. A system for document searching according to claim 20 and wherein said query is a question.

30 22. A system for document searching according to claim 20 and wherein said query is not a question.

23. A system for document searching according to any of claims 20 to 22 and wherein said computerized answer retrieving functionality is operative to provide said at least one additional search term, by retrieving search terms acquired other than in response to earlier questions, received by said computerized answer retrieving  
5 functionality prior to receipt of said query from said user.

24. A system for document searching according to any of claims 20 to 23 and wherein said computer is operative to receive said query from at least one of:  
a keyboard;  
10 a voice responsive input device;  
a screen scraping functionality;  
an email functionality;  
an SMS functionality; and  
an instant messaging functionality.

15 25. A system for document searching according to any of claims 21 to 24 and wherein said computerized answer retrieving functionality includes computerized query normalizing functionality for normalizing said query.

20 26. A system for document searching according to claim 25 and wherein said computerized query normalizing functionality is operative to normalize said query based at least in part on at least one of a plurality of query normalization rules.

25 27. A system for document searching according to any of claims 20 to 26 and wherein said computerized answer retrieving functionality is operative to generate said at least one additional search term not present in said query by replacing at least one word in said query by at least one selected synonym thereof.

30 28. A system for document searching according to claim 27 and wherein said computerized answer retrieving functionality includes computerized synonym retrieving functionality operative to identify said at least one selected synonym at least partially by

reference to at least one word in said query other than said at least one word which is replaced by said at least one selected synonym.

29. A system for document searching according to claim 28 and wherein said  
5 computerized synonym retrieving functionality includes a corpus and said computerized  
synonym retrieving functionality is operative to search said corpus for occurrences of at  
least one of a plurality of synonyms for which there exists a phrase relevant to said  
query and to designate at least one of said plurality of synonyms as a selected synonym  
in accordance with a number of occurrences in said corpus of a phrase including said at  
10 least one synonym which is relevant to said query.

30. A system for document searching according to any of claims 20 to 29  
and also comprising a document output device for providing a representation of at least  
one document in said set of documents to said user.

15

31. A system for document searching according to claim 30 and wherein said  
document output device comprises a display for presenting at least one link to said at  
least one document.

20 32. A system for document searching according to any of claims 20 to 31  
and also comprising:

computerized answer extraction functionality for extracting at least one  
answer from at least one document in said set of documents; and

25 an answer output device for providing said at least one answer to said  
user.

33. A system for document searching according to claim 32 and wherein said  
computerized answer extraction functionality includes a document analyzer operative to  
analyze said at least one document, said document analyzer including:

30 computerized theme extraction functionality for carrying out theme  
extraction on said at least one document, said theme extraction utilizing statistical

analysis of frequency of occurrence of words to identify at least one theme word of said at least one document;

computerized sentence extracting functionality for extracting sentences from said at least one document;

5 a potential answer selector for selecting at least one of said sentences as a potential answer;

computerized scoring functionality for scoring each of said at least one of said sentences; and

10 a sentence identifier for identifying at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

34. A system for document searching according to claim 32 and wherein said answer output device comprises a display for presenting said at least one answer to said user in an editable report precursor format.

15

35. A system for document searching according to any of claims 20 to 34 and wherein said computerized answer retrieving functionality includes artificial intelligence.

20

36. An answer extraction method comprising:  
employing a computer to receive a question from a user;  
employing a computer network to access a set of documents relevant to said question by employing document search terms derived by said computer from said question, said document search terms including at least one additional search term not present in the question, which said at least one additional search term was acquired prior to receipt of said question from said user;

25

analyzing said set of documents to extract at least one answer to said question; and

providing said at least one answer to said user.

30

37. An answer extraction method according to claim 36 and wherein said employing a computer network includes providing said at least one additional search



term, by retrieving search terms acquired in response to earlier questions, received prior to receipt of said question from said user.

38. An answer extraction method according to claim 36 and wherein said  
5 employing a computer network includes providing said at least one additional search  
term by retrieving search terms, acquired other than in response to earlier questions,  
received prior to receipt of said question from said user.

39. An answer extraction method according to either of claims 36 and 37 and  
10 wherein said employing a computer network employs artificial intelligence.

40. An answer extraction method according to any of claims 36 to 39 and  
wherein said employing a computer to receive a question comprises employing said  
computer to receive said question by at least one of:  
15 typing said question;  
using a voice responsive input device;  
using a screen scraping functionality;  
using an email functionality;  
using an SMS functionality; and  
20 using an instant messaging functionality.

41. An answer extraction method according to any of claims 36 to 40 and  
wherein said employing document search terms comprises utilizing computerized  
question normalizing functionality for normalizing said question.  
25

42. An answer extraction method according to claim 41 and wherein said  
normalizing said question is performed based at least in part on at least one of a  
plurality of question normalization rules.

30 43. An answer extraction method according to any of claims 36 to 42 and  
wherein said employing document search terms comprises generating document search

terms including said at least one additional search term not present in said question by replacing at least one word in said question by at least one selected synonym thereof.

44. An answer extraction method according to claim 43 and wherein said replacing at least one word in said question by at least one selected synonym thereof comprises employing computerized synonym retrieving functionality to identify said at least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected synonym.

45. An answer extraction method according to claim 44 and wherein said employing computerized synonym retrieving functionality comprises identifying said at least one selected synonym by:

identifying a plurality of synonyms; and

selecting at least one of said plurality of synonyms for which there exists a phrase relevant to said question in a corpus.

46. An answer extraction method according to claim 45 and wherein said identifying said at least one selected synonym comprises:

searching said corpus for occurrences of at least one of said plurality of synonyms for which there exists a phrase relevant to said question; and

designating at least one synonym of said plurality of synonyms as a selected synonym in accordance with a number of occurrences in said corpus of a phrase including said at least one synonym which is relevant to said question.

47. An answer extraction method according to any of claims 36 to 46 and also comprising utilizing computerized question processing functionality to process said question, said utilizing computerized question processing functionality including:

utilizing said computerized question processing functionality to generate at least one expected answer to said question;

utilizing said computerized question processing functionality to generate at least one preliminary search engine query based on said at least one expected answer;

utilizing said computerized question processing functionality to concatenate said at least one preliminary search engine query with said at least one additional search term not present in said question, thereby to form a concatenated search engine query; and

5                deriving said document search terms from said concatenated search engine query.

48.            An answer extraction method according to any of claims 36 to 47 and wherein said providing said at least one answer to said user also comprises providing a representation of at least one document of said set of documents to said user.

10

49.            An answer extraction method according to claim 48 and wherein said providing a representation comprises presenting at least one link to said at least one document.

15

50.            An answer extraction method according to any of claims 36 to 49 and wherein said analyzing said set of documents to extract at least one answer to said question comprises:

                 carrying out theme extraction on plural ones of said set of documents, said theme extraction utilizing statistical analysis of frequency of occurrence of words to identify at least one theme word of said at least one document;

20

                 extracting sentences from said at least one document;

                 selecting at least one of said sentences as a potential answer;

                 scoring each of said at least one of said sentences; and

25                identifying at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

51.            An answer extraction method according to any of claims 36 to 50 and wherein said analyzing said set of documents to extract said at least one answer comprises:

30

                 enhancing said at least one document of said set of documents by:

identifying capitalized phrases which appear in said at least one document;

identifying designated capitalized words belonging to said capitalized phrases; and

5 adding, to said at least one document adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, said designated capitalized word that does appear alongside thereof elsewhere in said at least one document in a capitalized phrase; and

10 carrying out analysis of said at least one document in order to identify at least one portion thereof as a potential answer.

52. An answer extraction method according to any of claims 36 to 51 and wherein said providing said at least one answer to said user comprises presenting said at least one answer in an editable report precursor format.

15

53. An answer extraction method according to any of claims 36 to 52 and wherein said question is not phrased in question format.

54. An answer extraction system comprising:  
20 a computer operative to receive a question from a user;  
computerized answer extraction functionality operative to employ a computer network to access a set of documents relevant to said question by employing document search terms derived by said computer from said question, said document search terms including at least one additional search term not present in the question,  
25 which said at least one additional search term was acquired prior to receipt of said question from said user;

computerized answer analysis functionality for analyzing said set of documents to extract at least one answer to said question; and  
an output device operative to provide said at least one answer to said  
30 user.

55. An answer extraction system according to claim 54 and wherein said computer network provides said at least one additional search term by retrieving search terms, acquired in response to earlier questions, received prior to receipt of said question from said user.

5

56. An answer extraction system according to claim 54 and wherein said computer network provides said at least one additional search term by retrieving search terms, acquired other than in response to earlier questions, received prior to receipt of said question from said user.

10

57. An answer extraction system according to any of claims 54 and 55 and wherein said computer network employs artificial intelligence.

58. An answer extraction system according to any of claims 54 to 57 and wherein said computer is operative to receive said question from at least one of:

15

- a keyboard;
- a voice responsive input device;
- a screen scraping functionality;
- an email functionality;
- an SMS functionality; and
- an instant messaging functionality.

20

59. An answer extraction system according to any of claims 54 to 58 and wherein said computerized answer extraction functionality includes computerized question normalizing functionality for normalizing said question.

25

60. An answer extraction system according to claim 59 and wherein said computerized question normalizing functionality is operative to normalize said question based at least in part on at least one of a plurality of question normalization rules.

30

61. An answer extraction system according to any of claims 54 to 60 and wherein said computerized answer extraction functionality is operative to generate said

at least one additional search term not present in said question by replacing at least one word in said question by at least one selected synonym thereof.

62. An answer extraction system according to claim 61 and wherein said  
5 computerized answer extraction functionality includes computerized synonym  
retrieving functionality operative to identify said at least one selected synonym at least  
partially by reference to at least one word in said question other than said at least one  
word which is replaced by said at least one selected synonym.

10 63. An answer extraction system according to claim 62 and wherein said  
computerized synonym retrieving functionality includes a corpus and said computerized  
synonym retrieving functionality is operative to search said corpus for occurrences of  
each one of a plurality of synonyms for which there exists a phrase including said one of  
said plurality of synonyms relevant to said question, and to designate at least one of said  
15 plurality of synonyms as a selected synonym in accordance with a number of  
occurrences in said corpus of a phrase including said at least one of said plurality of  
synonyms relevant to said question.

64. An answer extraction system according to any of claims 54 to 63 and  
20 wherein said output device is operative to provide a representation of at least one  
document of said set of documents to said user.

65. An answer extraction system according to claim 64 and wherein said  
output device comprises a display for presenting at least one link to said at least one  
25 document to said user.

66. An answer extraction system according to any of claims 54 to 65 and  
wherein said computerized answer extraction functionality includes:  
computerized theme extraction functionality for carrying out theme  
30 extraction on plural ones of said set of documents, said theme extraction utilizing  
statistical analysis of frequency of occurrence of words to identify at least one theme  
word of said at least one document;

computerized sentence extracting functionality for extracting sentences from said at least one document;

a potential answer selector for selecting at least one of said sentences as a potential answer;

5                   scoring functionality for scoring each said at least one of said sentences; and

a sentence identifier for identifying at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

10   67.           An answer extraction system according to any of claims 54 to 66 and wherein said output device comprises a display for presenting said at least one answer in an editable report precursor format.

15   68.           An answer extraction system according to any of claims 54 to 67 and wherein said question is not phrased in question format.

69.           An answer extraction method comprising:  
employing a computer to receive a question from a user;  
employing a computer network to access a set of documents relevant to  
20   said question by employing document search terms derived by said computer from said question;

extracting at least one answer to said question; and

providing said at least one answer to said user,

said extracting at least one answer comprising:

25                   generating an expected answer to said question, said expected answer including question keywords;

analyzing said set of documents by:

30                   carrying out theme extraction on plural ones of said set of documents, said theme extraction utilizing statistical analysis of the frequency of occurrence of words to identify at least one theme word of a document, which theme word may or may not be a question keyword; and

extracting sentences from plural ones of said set of documents;

selecting at least one of said sentences as a potential answer if it fulfills at least one of the following criteria:

5 a sentence including at least a predetermined plurality of question keywords; and

a sentence including at least one question keyword and at least one theme word;

10 scoring each of said at least one of said sentences selected as a potential answer; and

identifying at least one of said at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

15 70. An answer extraction method according to claim 69 and wherein said employing a computer to receive a question comprises employing said computer to receive said question by at least one of:

typing said question;  
using a voice responsive input device;  
20 using a screen scraping functionality;  
using an email functionality;  
using an SMS functionality; and  
using an instant messaging functionality.

25 71. An answer extraction method according to either of claims 69 and 70 and also comprising, prior to said employing a computer network to access a set of documents:

utilizing computerized question normalization functionality for normalizing said question; and  
30 thereafter, utilizing computerized question classification functionality to classify said question.



72. An answer extraction method according to claim 71 and wherein said normalizing said question is performed based at least in part on at least one of a plurality of question normalization rules.

5 73. An answer extraction method according to any of claims 69 to 72 and wherein said employing a computer network comprises employing said computer to derive said document search terms including at least one additional search term not present in the question, which said at least one additional search term was acquired prior to receipt of said question from said user.

10

74. An answer extraction method according to any of claims 69 to 72 and wherein said employing a computer network comprises employing said computer to derive said document search terms including at least one additional search term not present in the question by replacing at least one word in said question by at least one  
15 selected synonym thereof.

75. An answer extraction method according to claim 74 and wherein said replacing at least one word in said question by at least one selected synonym thereof comprises employing computerized synonym retrieving functionality to identify said at  
20 least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected synonym.

76. An answer extraction method according to claim 75 and wherein said  
25 employing computerized synonym retrieving functionality comprises identifying said at least one selected synonym by:

identifying a plurality of synonyms; and

selecting at least one of said plurality of synonyms for which there exists a phrase relevant to said question in a corpus.

30

77. An answer extraction method according to claim 76 and wherein said identifying said at least one selected synonym comprises:

searching said corpus for occurrences of at least one of said plurality of synonyms for which there exists a phrase relevant to said question; and

designating at least one of said plurality of synonyms as a selected synonym in accordance with a number of occurrences in said corpus of a phrase including said at least one of said plurality of synonyms which is relevant to said question.

78. An answer extraction method according to any of claims 69 to 77 and also comprising providing a representation of said at least one document of said set of documents to said user.

79. An answer extraction method according to claim 78 and wherein said providing a representation comprises presenting at least one link to said at least one document.

80. An answer extraction method according to any of claims 69 to 79 and wherein said providing said at least one answer to said user comprises presenting said at least one answer in an editable report precursor format.

81. An answer extraction method according to any of claims 69 to 80 and wherein said statistical analysis comprises:

for each word in said document, stemming said word to a corresponding root word;

generating a word occurrence frequency score for each different root word corresponding to a word in said document;

using said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said document;

selecting a subset of words in said document including at least one word having a word occurrence frequency score which is greater than or equal to said document word occurrence frequency indicating score.

82. An answer extraction method according to claim 81 and wherein said document word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

5

83. An answer extraction method according to either of claims 81 and 82 and wherein said statistical analysis comprises selecting, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

10

84. An answer extraction method according to any of claims 81 to 83 and wherein said statistical analysis also comprises:

following said selecting a subset of words in said document, calculating a subset word occurrence frequency indicating score; and

15

selecting, as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

85. An answer extraction method according to claim 84 and wherein said subset word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

20

86. An answer extraction method according to any of claims 69 to 85 and wherein said question is not phrased in question format.

25

87. An answer extraction system comprising:

a computer operative to receive a question from a user; and

computerized answer extraction functionality operative to employ a

computer network to access a set of documents relevant to said question by employing document search terms derived by said computer from said question, to extract at least one answer to said question and to provide said at least one answer to said user,

30

said computerized answer extraction functionality comprising:

an expected answer generator operative to generate an expected answer to said question, said expected answer including question keywords;

5 a document analyzer operative to carry out theme extraction on plural ones of said set of documents, said theme extraction utilizing statistical analysis of the frequency of occurrence of words in a document to identify at least one theme word of said document, which theme word may or may not be a question keyword;

a sentence extractor, operative to extract sentences from plural ones of said set of documents;

10 a potential answer selector, operative to select at least one of said sentences as a potential answer if it fulfills at least one of the following criteria:

a sentence including at least a predetermined plurality of question keywords; and

15 a sentence including at least one question keyword and at least one theme word; and

a potential answer identifier, operative to calculate a score for each of said at least one of said sentences selected as a potential answer and to identify at least one of said sentences selected as a potential answer based at least partially on said score.

20

88. An answer extraction system according to claim 87 and wherein said computer is operative to receive said question from at least one of:

a keyboard;

a voice responsive input device;

25 a screen scraping functionality;

an email functionality;

an SMS functionality; and

an instant messaging functionality.

30 89. An answer extraction system according to either of claims 87 and 88 and also comprising:

computerized question normalizing functionality operative to normalize said question; and

computerized question classification functionality for classifying said question.

5

90. An answer extraction system according to claim 89 and wherein said computerized question normalizing functionality is operative to normalize said question based at least in part on at least one of a plurality of question normalization rules.

10 91. An answer extraction system according to any of claims 87 to 90 and wherein said computerized answer extraction functionality is operative to employ said computer to derive said document search terms, including at least one additional search term not present in the question, which said at least one additional search term was acquired prior to receipt of said question from said user.

15

92. An answer extraction system according to any of claims 87 to 90 and wherein said computerized answer extraction functionality is operative to employ said computer to derive said document search terms, including at least one additional search term not present in the question, by replacing at least one word in said question by at least one selected synonym thereof.

20

93. An answer extraction system according to claim 92 and wherein said computerized answer extraction functionality includes computerized synonym retrieving functionality operative to identify said at least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected synonym.

25

94. An answer extraction system according to claim 93 and wherein said computerized synonym retrieving functionality includes a corpus and said computerized synonym retrieving functionality is operative to search said corpus for occurrences of at least one of a plurality of synonyms for which there exists a phrase relevant to said question and to designate at least one synonym as a selected synonym in accordance

30

with a number of occurrences in said corpus of a phrase, including said at least one synonym which is relevant to said question.

5 95. An answer extraction system according to any of claims 87 to 94 and also comprising a document output device for providing a representation of at least one document of said set of documents to said user.

10 96. An answer extraction system according to claim 95 and wherein said document output device comprises a display for presenting at least one link to said at least one document.

15 97. An answer extraction system according to any of claims 87 to 96 and also comprising an answer output device for providing said at least one answer to said user.

98. An answer extraction system according to claim 97 and wherein said answer output device comprises a display for presenting said at least one answer in an editable report precursor format.

20 99. An answer extraction system according to any of claims 87 to 98 and wherein said document analyzer comprises:

computerized word stemming functionality, operative, for each word in said document, to stem said word to a corresponding root word;

25 a word occurrence frequency score generator for generating a word occurrence frequency score for each different root word corresponding to a word in said document;

30 computerized document word occurrence frequency indicating score calculating functionality operative to use said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said document; and

computerized word selecting functionality operative to select a subset of words in said document including at least one word having a word occurrence

frequency score which is greater than or equal to said document word occurrence frequency indicating score.

100. An answer extraction system according to claim 99 and wherein said  
5 computerized document word occurrence frequency indicating score calculating functionality is operative to calculate said document word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

10 101. An answer extraction system according to either of claims 99 and 100 and wherein said computerized word selecting functionality is operative to select, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

15 102. An answer extraction system according to any of claims 99 to 101 and wherein said document analyzer also comprises:

computerized subset word occurrence frequency indicating score  
calculating functionality, operative to calculate a subset word occurrence frequency  
20 indicating score; and

computerized theme word selection functionality operative to select, as  
said at least one theme word, at least one of said subset of words having a word  
occurrence frequency score which is greater than or equal to said subset word  
occurrence frequency indicating score.

25 103. An answer extraction system according to claim 102 and wherein said computerized subset word occurrence frequency indicating score calculating functionality is operative to calculate said subset word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores  
30 of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

104. An answer extraction system according to any of claims 87 to 103 and wherein said question is not phrased in question format.

105. An answer extraction method comprising:

5 employing a computer to receive a question from a user;  
employing a computer network to access a set of documents relevant to said question by employing document search terms derived by said computer from said question;

10 extracting at least one answer to said question; and

providing said at least one answer to said user,

said extracting at least one answer including:

enhancing at least one of said set of documents by:

identifying capitalized phrases which appear in said at least one document;

15 identifying designated capitalized words belonging to said capitalized phrases; and

adding, to said at least one document adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, the designated capitalized word that does appear alongside thereof elsewhere in the document in a capitalized phrase; and

20 carrying out analysis of said at least one document in order to identify at least one portion thereof as a potential answer.

106. An answer extraction method according to claim 105 and wherein said employing a computer to receive a question comprises employing said computer to receive said question by at least one of:

typing said question;  
using a voice responsive input device;  
using a screen scraping functionality;  
30 using an email functionality;  
using an SMS functionality; and  
using an instant messaging functionality.



107. An answer extraction method according to either of claims 105 and 106 and also comprising, prior to said employing said computer network:

utilizing computerized question normalization functionality for  
5 normalizing said question; and

thereafter, utilizing computerized question classification functionality to classify said question.

108. An answer extraction method according to claim 107 and wherein said  
10 normalizing said question is performed based at least in part on at least one of a plurality of question normalization rules.

109. An answer extraction method according to any of claims 105 to 108 and wherein said employing a computer network comprises employing said computer to  
15 derive said document search terms, including at least one additional search term not present in the question, which said at least one additional search term was acquired prior to receipt of said question from said user.

110. An answer extraction method according to any of claims 105 to 109 and  
20 wherein said employing a computer network comprises employing said computer to derive said document search terms, including at least one additional search term not present in the question by replacing at least one word in said question by at least one selected synonym thereof.

25 111. An answer extraction method according to claim 110 and wherein said replacing at least one word in said question by at least one selected synonym thereof comprises employing computerized synonym retrieving functionality to identify said at least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected  
30 synonym.

112. An answer extraction method according to claim 111 and wherein said employing computerized synonym retrieving functionality comprises identifying said at least one selected synonym by:

- identifying a plurality of synonyms; and
- 5 selecting at least one of said plurality of synonyms for which there exists a phrase relevant to said question in a corpus.

113. An answer extraction method according to claim 112 and wherein said identifying said at least one selected synonym comprises:

- 10 searching said corpus for occurrences of at least one of said plurality of synonyms for which there exists a phrase relevant to said question; and
- designating at least one synonym as a selected synonym in accordance with a number of occurrences in said corpus of a phrase including said at least one synonym which is relevant to said question.

15

114. An answer extraction method according to any of claims 105 to 113 and wherein said extracting at least one answer also comprises, prior to said enhancing, generating an expected answer to said question, said expected answer including question keywords, and wherein said carrying out analysis of said at least one document

20 comprises:

- carrying out theme extraction on said at least one document, said theme extraction utilizing statistical analysis of the frequency of occurrence of words to identify at least one theme word of said at least one document, which theme word may or may not be a question keyword;
- 25 extracting sentences from said at least one document;
- selecting at least one of said sentences as a potential answer if it fulfills at least one of the following criteria:
  - a sentence including at least a predetermined plurality of question keywords; and
  - 30 a sentence including at least one question keyword and at least one theme word;

scoring each of said at least one of said sentences selected as a potential answer; and

identifying at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

5

115. An answer extraction method according to claim 114 and wherein said statistical analysis comprises:

for each word in said at least one document, stemming said word to a corresponding root word;

10 generating a word occurrence frequency score for each different root word corresponding to a word in said at least one document;

using said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said at least one document; and

15 selecting as potential theme words a subset of words in said at least one document including at least one word having a word occurrence frequency score which is greater than or equal to said document word occurrence frequency indicating score.

116. An answer extraction method according to claim 115 and wherein said document word occurrence frequency indicating score comprises at least one of an  
20 average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

117. An answer extraction method according to either of claims 115 and 116 and wherein said selecting as potential theme words comprises selecting, as said at least  
25 one theme word, at least one word having a word occurrence frequency score which greater than or equal to twice said document word occurrence frequency indicating score.

118. An answer extraction method according to any of claims 114 to 117 and  
30 wherein said statistical analysis also comprises:

following said selecting as potential theme words a subset of words in said at least one document, calculating a subset word occurrence frequency indicating score; and

5 selecting, as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

119. An answer extraction method according to claim 118 and wherein said subset word occurrence frequency indicating score comprises at least one of an average  
10 of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

120. An answer extraction method according to any of claims 105 to 119 and also comprising providing a representation of at least one of said set of documents to  
15 said user.

121. An answer extraction method according to claim 120 and wherein said providing a representation comprises presenting at least one link to said at least one document.

20

122. An answer extraction method according to any of claims 105 to 121 and wherein said providing said at least one answer to said user comprises presenting said at least one answer in an editable report precursor format.

25 123. An answer extraction method according to any of claims 105 to 122 and wherein said question is not phrased in question format.

124. An answer extraction system comprising:  
a computer operative to receive a question from a user;  
30 computerized answer extraction functionality operative to employ a computer network to access a set of documents relevant to said question by employing

document search terms derived by said computer from said question, to extract at least one answer to said question and to provide said at least one answer to said user,

5                   said computerized answer extraction functionality comprising a document analyzer operative to identify capitalized phrases which appear in a document belonging to said set of documents, to identify designated capitalized words belonging to said capitalized phrases, to add to said document adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, the designated capitalized word that does appear alongside thereof elsewhere in said document in a capitalized phrase, thereby providing an enhanced document, and to carry out analysis  
10 of said enhanced document in order to identify at least one portion thereof as a potential answer.

125.           An answer extraction system according to claim 124 and wherein said computer is operative to receive said question from at least one of:

15                   a keyboard;  
                  a voice responsive input device;  
                  a screen scraping functionality;  
                  an email functionality;  
                  an SMS functionality; and  
20                   an instant messaging functionality.

126.           An answer extraction system according to either of claims 124 and 125 and also comprising:

                  computerized question normalizing functionality operative for  
25 normalizing said question; and  
                  computerized question classification functionality for classifying said question.

127.           An answer extraction system according to claim 126 and wherein said  
30 computerized question normalizing functionality is operative to normalize said question based at least in part on at least one of a plurality of question normalization rules.

128. An answer extraction system according to any of claims 124 to 127 and wherein said computerized answer extraction functionality is operative to employ said computer to derive said document search terms, including at least one additional search term not present in the question, which said at least one additional search term was  
5 acquired prior to receipt of said question from said user.

129. An answer extraction system according to any of claims 124 to 128 and wherein said computerized answer extraction functionality is operative to employ said computer to derive said document search terms, including at least one additional search  
10 term not present in the question by replacing at least one word in said question by at least one selected synonym thereof.

130. An answer extraction system according to claim 129 and wherein said computerized answer extraction functionality includes computerized synonym  
15 retrieving functionality operative to identify said at least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected synonym.

131. An answer extraction system according to claim 130 and wherein said  
20 computerized synonym retrieving functionality includes a corpus and said computerized synonym retrieving functionality is operative to search said corpus for occurrences of at least one of a plurality of synonyms for which there exists a phrase relevant to said question and to designate at least one of said plurality of synonyms as a selected synonym in accordance with a number of occurrences in said corpus of a phrase  
25 including said at least one of said plurality of synonyms which is relevant to said question.

132. An answer extraction system according to any of claims 124 to 131 and wherein said computerized answer extraction functionality also comprises an expected  
30 answer generator operative to generate an expected answer to said question, said expected answer including question keywords, and wherein said document analyzer comprises:

computerized theme extraction functionality for carrying out theme extraction on said document, said theme extraction utilizing statistical analysis of the frequency of occurrence of words to identify at least one theme word of said document, which theme word may or may not be a question keyword;

5 a sentence extractor, operative to extract sentences from said document;

a potential answer selector, operative to select at least one of said sentences as a potential answer if it fulfills at least one of the following criteria:

a sentence including at least a predetermined plurality of question keywords; and

10 a sentence including at least one question keyword and at least one theme word; and

a potential answer identifier, operative to calculate a score for each of said at least one of said sentences and to identify at least one of said sentences selected as a potential answer based at least partially on results of said score.

15

133. An answer extraction system according to claim 132 and wherein said document analyzer comprises:

computerized word stemming functionality, operative, for each word in said document, to stem said word to a corresponding root word;

20 a word occurrence frequency score generator for generating a word occurrence frequency score for each different root word corresponding to a word in said document;

computerized document word occurrence frequency indicating score calculating functionality operative to use said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said document; and

25 computerized word selecting functionality operative to select a subset of words in said document including at least one word having a word occurrence frequency score which is greater than or equal to said document word occurrence frequency indicating score.

30

134. An answer extraction system according to claim 133 and wherein said computerized document word occurrence frequency indicating score calculating functionality is operative to calculate said document word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

135. An answer extraction system according to any of claims 132 to 134 and wherein said computerized theme extraction functionality is operative to select, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

136. An answer extraction system according to any of claims 132 to 135 and wherein said document analyzer also comprises:

computerized subset word occurrence frequency indicating score calculating functionality, operative to calculate a subset word occurrence frequency indicating score; and

computerized theme word selection functionality operative to select, as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

137. An answer extraction system according to claim 136 and wherein said computerized subset word occurrence frequency indicating score calculating functionality is operative to calculate said subset word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

138. An answer extraction system according to any of claims 124 to 137 and also comprising a document output device for providing a representation of at least one of said set of documents to said user.



139. An answer extraction system according to claim 138 and wherein said document output device comprises a display for presenting at least one link to said at least one document.

5

140. An answer extraction system according to any of claims 124 to 139 and also comprising an answer output device for providing said at least one answer to said user.

10

141. An answer extraction system according to claim 140 and wherein said answer output device comprises a display for presenting said at least one answer in an editable report precursor format.

15

142. An answer extraction system according to any of claims 124 to 141 and wherein said question is not phrased in question format.

20

143. An answer extraction method comprising:  
employing a computer to receive a question from a user;  
employing a computer network to access a set of documents relevant to  
said question by employing document search terms derived by said computer from said question;

25

extracting at least one answer to said question; and  
providing said at least one answer to said user,  
said extracting at least one answer to said question comprising:  
identifying a multiplicity of potential answers; and  
evaluating each of said multiplicity of potential answers  
according to at least one of the following criteria:

30

proximity of question keywords in the potential answer;  
proximity of classification words and nouns in the  
potential answer; and  
word count of at least part of the potential answer.

144. An answer extraction method according to claim 143 and wherein said evaluating comprises evaluating each of said multiplicity of potential answers according to at least two of the following criteria:

- 5                   proximity of question keywords in the potential answer;  
                  proximity of classification words and nouns in the potential answer; and  
                  word count of at least part of the potential answer.

145. An answer extraction method according to claim 143 and wherein said evaluating comprises evaluating each of said multiplicity of potential answers according to all of the following criteria:

- 10                   proximity of question keywords in the potential answer;  
                  proximity of classification words and nouns in the potential answer; and  
                  word count of at least part of the potential answer.

146. An answer extraction method according to claim 143 and wherein said evaluating comprises evaluating each of said multiplicity of potential answers according to a combination of the following criteria:

- 15                   proximity of question keywords in the potential answer;  
                  proximity of classification words and nouns in the potential answer; and  
20                   word count of at least part of the potential answer.

147. An answer extraction method according to any of claims 143 to 146 and wherein said extracting at least one answer also comprises selecting a sub group of said multiplicity of potential answers based on an evaluation of said multiplicity of potential answers in accordance with said criteria.

148. An answer extraction method according to claim 147 and wherein said evaluation comprises scoring said multiplicity of potential answers in accordance with said criteria.

149. An answer extraction method according to claim 148 and also comprising:

forming a potential answer document by combining said multiplicity of potential answers;

extracting a theme of said sub group of said multiplicity of potential answers, by utilizing statistical analysis of the frequency of occurrence of words in said potential answer document to identify at least one theme word in said sub group of said multiplicity of potential answers, which theme word may or may not be a question keyword; and

discarding potential answers belonging to said sub group of said multiplicity of potential answers which do not include at least one of said at least one theme word.

150. An answer extraction method according to claim 149 and wherein said statistical analysis comprises:

for each word in said potential answer document, stemming said word to a corresponding root word;

generating a word occurrence frequency score for each different root word corresponding to a word in said potential answer document;

using said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said potential answer document; and

selecting a subset of words in said potential answer document including at least one word having a word occurrence frequency score which is greater than or equal to said document word occurrence frequency indicating score.

151. An answer extraction method according to claim 150 and wherein said document word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

152. An answer extraction method according to any of claims 149 to 151 and wherein said extracting a theme comprises selecting, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

153. An answer extraction method according to any of claims 149 to 152 and wherein said statistical analysis also comprises, following said selecting a subset of words in said potential answer document:

- 5                   calculating a subset word occurrence frequency indicating score; and  
                  selecting as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

- 10 154. An answer extraction method according to claim 153 and wherein said subset word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

- 15 155. An answer extraction method according to any of claims 143 to 154 and wherein said providing said at least one answer to said user comprises providing said at least one answer to said user in an order governed at least in part by at least one of:

- a word count of each of said at least one answer;  
                  a score resulting from application to each of said at least one answer of at  
20 least one of the following criteria:  
                  proximity of question keywords in said at least one answer;  
                  proximity of classification words and nouns in said at least one  
answer; and  
                  word count of at least part of said at least one answer.

- 25 156. An answer extraction method according to any of claims 143 to 155 and wherein said employing a computer to receive said question comprises employing said computer to receive said question by at least one of:

- typing said question;  
30                   using a voice responsive input device;  
                  using a screen scraping functionality;  
                  using an email functionality;

using an SMS functionality; and  
using an instant messaging functionality.

157. An answer extraction method according to any of claims 143 to 156 and  
5 also comprising, prior to said employing a computer network:

utilizing computerized question normalization functionality for  
normalizing said question; and

thereafter, utilizing computerized question classification functionality to  
classify said question.

10

158. An answer extraction method according to claim 157 and wherein said  
normalizing said question is performed based at least in part on at least one of a  
plurality of question normalization rules.

159. An answer extraction method according to any of claims 143 to 158 and  
wherein said employing a computer network comprises employing said computer to  
derive said document search terms, including at least one additional search term not  
present in the question, which said at least one additional search term was acquired prior  
to receipt of said question from said user.

20

160. An answer extraction method according to any of claims 143 to 159 and  
wherein said employing a computer network comprises employing said computer to  
derive said document search terms, including at least one additional search term not  
present in the question by replacing at least one word in said question by at least one  
25 selected synonym thereof.

161. An answer extraction method according to claim 160 and wherein said  
replacing at least one word in said question by at least one selected synonym thereof  
comprises employing computerized synonym retrieving functionality to identify said at  
30 least one selected synonym at least partially by reference to at least one word in said  
question other than said at least one word which is replaced by said at least one selected  
synonym.

162. An answer extraction method according to claim 161 and wherein said employing computerized synonym retrieving functionality comprises identifying said at least one selected synonym by:

- 5 identifying a plurality of synonyms; and  
selecting at least one of said plurality of synonyms for which there exists a phrase relevant to said question in a corpus.

163. An answer extraction method according to claim 162 and wherein said identifying said at least one selected synonym comprises:

- searching said corpus for occurrences of at least one of said plurality of synonyms for which there exists a phrase relevant to said question; and  
designating at least one of said plurality of synonyms as a selected synonym in accordance with a number of occurrences in said corpus of a phrase  
15 including said at least one of said plurality of synonyms which is relevant to said question.

164. An answer extraction method according to any of claims 143 to 163 and wherein said identifying a multiplicity of potential answers also comprises:

- 20 enhancing at least one of said set of documents by:  
identifying capitalized phrases which appear in said at least one of said set of documents;  
identifying designated capitalized words belonging to said capitalized phrases; and  
25 adding, to said at least one of said set of documents adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, the designated capitalized word that does appear alongside thereof elsewhere in the document in a capitalized phrase; and  
carrying out analysis of said at least one of said set of documents in order  
30 to identify at least one portion thereof as a potential answer.

165. An answer extraction method according to claim 164 and wherein said identifying a multiplicity of potential answers also comprises, prior to said enhancing, generating an expected answer to said question, said expected answer including question keywords, and wherein said carrying out analysis comprises:

5 carrying out theme extraction on said at least one of said set of documents, said theme extraction utilizing statistical analysis of the frequency of occurrence of words to identify at least one theme word of said at least one of said set of documents, which theme word may or may not be a question keyword;

extracting sentences from said at least one of said set of documents;

10 selecting at least one of said sentences as a potential answer if it fulfills at least one of the following criteria:

a sentence including at least a predetermined plurality of question keywords; and

15 a sentence including at least one question keyword and at least one theme word;

scoring each of said at least one of said sentences selected as a potential answer; and

identifying at least one of said sentences selected as a potential answer based at least partially on results of said scoring.

20

166. An answer extraction method according to any of claims 143 to 165 and also comprising providing a representation of at least one of said set of documents to said user.

25 167. An answer extraction method according to claim 166 and wherein said providing a representation comprises presenting at least one link to said at least one of said set of documents.

30 168. An answer extraction method according to any of claims 143 to 167 and wherein said providing said at least one answer to said user comprises presenting said at least one answer in an editable report precursor format.

169. An answer extraction method according to any of claims 143 to 168 and wherein said question is not phrased in question format.

170. An answer extraction system comprising:

5 a computer operative to receive a question from a user;

computerized answer extraction functionality operative to employ a computer network to access a set of documents relevant to said question by employing document search terms derived by said computer from said question, to extract at least one answer to said question and to provide said at least one answer to said user, said  
10 computerized answer extraction functionality being operative to identify a multiplicity of potential answers and to evaluate each of said multiplicity of potential answers according to at least one of the following criteria:

proximity of question keywords in the potential answer;

proximity of classification words and nouns in the potential  
15 answer; and

word count of at least part of the potential answer.

171. An answer extraction system according to claim 170 and wherein said computerized answer extraction functionality is operative to evaluate each of said  
20 multiplicity of potential answers according to at least two of the following criteria:

proximity of question keywords in the potential answer;

proximity of classification words and nouns in the potential answer; and

word count of at least part of the potential answer.

25 172. An answer extraction system according to claim 171 and wherein said computerized answer extraction functionality is operative to evaluate each of said multiplicity of potential answers according to all of the following criteria:

proximity of question keywords in the potential answer;

proximity of classification words and nouns in the potential answer; and

30 word count of at least part of the potential answer.



173. An answer extraction system according to claim 172 and wherein said computerized answer extraction functionality is operative to evaluate each of said multiplicity of potential answers according to a combination of the following criteria:

- 5                   proximity of question keywords in the potential answer;  
                  proximity of classification words and nouns in the potential answer; and  
                  word count of at least part of the potential answer.

174. An answer extraction system according to any of claims 170 to 173 and wherein said computerized answer extraction functionality is also operative to select a  
10 sub group of said multiplicity of potential answers based on an evaluation of said multiplicity of potential answers in accordance with said criteria.

175. An answer extraction system according to claim 174 and wherein said evaluation comprises scoring said multiplicity of potential answers in accordance with  
15 said criteria.

176. An answer extraction system according to claim 175 and also comprising:

20                   computerized potential answer combining functionality operative to form a potential answer document by combining said multiplicity of potential answers;

                  computerized theme extraction functionality for carrying out theme extraction on said sub group of said multiplicity of potential answers, said theme extraction utilizing statistical analysis of the frequency of occurrence of words in said potential answer document to identify at least one theme word in said sub group of said  
25 multiplicity of potential answers, which theme word may or may not be a question keyword; and

                  computerized potential answer discarding functionality operative to discard potential answers belonging to said sub group of said multiplicity of potential answers which do not include at least one of said at least one theme word.

30

177. An answer extraction system according to claim 176 and wherein said computerized theme extraction functionality comprises:

computerized word stemming functionality, operative, for each word in said potential answers document, to stem said word to a corresponding root word;

5 a word occurrence frequency score generator for generating a word occurrence frequency score for each different root word corresponding to a word in said potential answers document;

computerized document word occurrence frequency indicating score calculating functionality operative to use said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said potential answers document; and

10 computerized word selecting functionality operative to select a subset of words in said potential answers document including at least one word having a word occurrence frequency score which is greater than or equal to said document word occurrence frequency indicating score.

15 178. An answer extraction system according to claim 177 and wherein said computerized document word occurrence frequency indicating score calculating functionality is operative to calculate said document word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

20 179. An answer extraction system according to any of claims 176 to 178 and wherein said computerized theme extraction functionality is operative to select, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency  
25 indicating score.

180. An answer extraction system according to any of claims 176 to 179 and also comprising:

30 computerized subset word occurrence frequency indicating score calculating functionality, operative to calculate a subset word occurrence frequency indicating score; and

computerized theme word selection functionality operative to select, as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

5

181. An answer extraction system according to claim 180 and wherein said computerized subset word occurrence frequency indicating score calculating functionality is operative to calculate said subset word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

10

182. An answer extraction system according to any of claims 170 to 181 and wherein said computerized answer extraction functionality provides said at least one answer to said user in an order governed at least in part by at least one of:

15

a word count of each one of said at least one answer; and

a score, resulting from application to each one of said at least one answer of at least one of the following criteria:

20

proximity of question keywords in said at least one answer;

proximity of classification words and nouns in said at least one answer; and

word count of at least part of said at least one answer.

183. An answer extraction system according to any of claims 170 to 182 and wherein said computer is operative to receive said question from at least one of:

25

a keyboard;

a voice responsive input device;

a screen scraping functionality;

an email functionality;

30

an SMS functionality; and

an instant messaging functionality.

184. An answer extraction system according to any of claims 170 to 183 and also comprising:

computerized question normalizing functionality for normalizing said question; and

5 computerized question classification functionality for classifying said question.

185. An answer extraction system according to claim 184 and wherein said computerized question normalizing functionality is operative to normalize said question  
10 based at least in part on at least one of a plurality of question normalization rules.

186. An answer extraction system according to any of claims 170 to 185 and wherein said computerized answer extraction functionality is operative to employ said computer to derive said document search terms, including at least one additional search  
15 term not present in the question, which said at least one additional search term was acquired prior to receipt of said question from said user.

187. An answer extraction system according to any of claims 170 to 185 and wherein said computerized answer extraction functionality is operative to employ said  
20 computer to derive said document search terms, including at least one additional search term not present in the question by replacing at least one word in said question by at least one selected synonym thereof.

188. An answer extraction system according to claim 187 and wherein said  
25 computerized answer extraction functionality includes computerized synonym retrieving functionality operative to identify said at least one selected synonym at least partially by reference to at least one word in said question other than said at least one word which is replaced by said at least one selected synonym.

30 189. An answer extraction system according to claim 188 and wherein said computerized synonym retrieving functionality includes a corpus and said computerized synonym retrieving functionality is operative to search said corpus for occurrences of at

least one of a plurality of synonyms for which there exists a phrase relevant to said question and to designate at least one of said plurality of synonyms as a selected synonym in accordance with a number of occurrences in said corpus of a phrase, including said at least one of said plurality of synonyms relevant to said question.

5

190. An answer extraction system according to any of claims 170 to 189 and wherein said computerized answer extraction functionality comprises computerized document analysis functionality operative to identify capitalized phrases which appear in at least one of said set of documents, to identify designated capitalized words  
10 belonging to said capitalized phrases and to add to said at least one of said set of documents, adjacent each occurrence of a designated capitalized word that does not appear in a capitalized phrase, the designated capitalized word that does appear alongside thereof elsewhere in said at least one of said set of documents in a capitalized phrase, thereby providing an enhanced document, and to carry out analysis of said  
15 enhanced document in order to identify at least one portion thereof as a potential answer.

191. An answer extraction system according to claim 190 and wherein said computerized answer extraction functionality also comprises an expected answer  
20 generator operative to generate an expected answer to said question, said expected answer including question keywords, and wherein said computerized document analysis functionality comprises:

computerized theme extraction functionality for carrying out theme extraction on said enhanced document, said theme extraction utilizing statistical  
25 analysis of the frequency of occurrence of words to identify at least one theme word of said enhanced document, which theme word may or may not be a question keyword;

a sentence extractor, operative to extract sentences from said enhanced document;

a potential answer selector, operative to select at least one of said  
30 sentences as a potential answer if it fulfills at least one of the following criteria:

a sentence including at least a predetermined plurality of question keywords; and

a sentence including at least one question keyword and at least one theme word; and

5 a potential answer identifier, operative to calculate a score for each of said at least one of said sentences selected as a potential answer and to identify at least one of said sentences selected as a potential answer based at least partially on results of said score.

192. An answer extraction system according to any of claims 170 to 191 and also comprising a document output device for providing a representation of at least one  
10 of said set of documents to said user.

193. An answer extraction system according to claim 192 and wherein said document output device comprises a display for presenting at least one link to said at least one of said set of documents.

15 194. An answer extraction system according to any of claims 170 to 193 and also comprising an answer output device for providing said at least one answer to said user.

20 195. An answer extraction system according to claim 194 and wherein said answer output device comprises a display for presenting said at least one answer in an editable report precursor format.

25 196. An answer extraction method according to any of claims 170 to 195 and wherein said question is not phrased in question format.

197. A document searching method comprising:  
employing a computer to receive a query including at least one search term from a user; and  
30 employing computerized synonym retrieving functionality operative in response to queries to generate document search terms including at least one additional search term not present in said query, said computerized synonym retrieving

functionality being operative to generate said at least one additional search term by replacing at least one word in said query by at least one selected synonym thereof; and

operating computerized search engine functionality to access a set of documents in response to said query, based on at least one of said at least one search  
5 term supplied by a user and said at least one additional search term provided by said computerized synonym retrieving functionality,

said computerized synonym retrieving functionality being operative to identify said at least one selected synonym at least partially by reference to at least one word in said query other than said at least one word.

10

198. A document searching method according to claim 197 and wherein said computerized synonym retrieving functionality is operative to identify said at least one selected synonym by:

identifying a plurality of synonyms; and

15

selecting at least one of said plurality of synonyms for which there exists a phrase relevant to said query in a corpus.

199. A document searching method according to claim 198 and wherein said computerized synonym retrieving functionality is operative to identify said selected  
20 synonym by:

searching said corpus for occurrences of said at least one of said plurality of synonyms for which there exists a phrase relevant to said query; and

designating at least one of said plurality of synonyms as a selected synonym in accordance with the number of occurrences in said corpus of a phrase  
25 including said at least one of said plurality of synonyms which is relevant to said query.

200. A document searching method according to any of claims 197 to 199 and wherein said query is a question.

30 201. A document searching method according to any of claims 197 to 199 and wherein said query is not a question.

202. A document searching method according to any of claims 197 to 201 and wherein said at least one word in said query which is replaced by said at least one selected synonym thereof comprises at least one of a noun, a verb, an object of a verb and a subject of a verb.

5

203. A document searching system comprising:  
a computer operative to receive a query including at least one search term from a user;

computerized synonym retrieving functionality operative, in response to queries, to generate document search terms, including at least one additional search term not present in said query and to generate said at least one additional search term by replacing at least one word in said query by at least one selected synonym thereof; and

computerized search engine functionality operative to access a set of documents in response to said query, based on at least one of said at least one search term supplied by a user and said at least one additional search term provided by said computerized synonym retrieving functionality,

said computerized synonym retrieving functionality being operative to identify said selected synonym at least partially by reference to a word in said query other than said at least one word.

20

204. A document searching system according to claim 203 and wherein said computerized synonym retrieving functionality comprises a synonym selector operative to identify a plurality of synonyms and to select at least one of said plurality of synonyms for which there exists a phrase relevant to said query in a corpus.

25

205. A document searching system according to claim 204 and wherein said synonym selector is operative to identify said selected synonym by:

searching said corpus for occurrences of said at least one of said plurality of synonyms for which there exists a phrase relevant to said query; and

designating at least one of said plurality of synonyms as a selected synonym in accordance with the number of occurrences in said corpus of a phrase including said at least one of said plurality of synonyms which is relevant to said query.

30



206. A document searching system according to any of claims 203 to 205 and wherein said query is a question.
- 5 207. A document searching system according to any of claims 203 to 205 and wherein said query is not a question.
208. A document searching system according to any of claims 203 to 207 and wherein said at least one word in said query which is replaced by said at least one  
10 selected synonym thereof comprises at least one of a noun, a verb, an object of a verb and a subject of a verb.
209. A computerized synonym generating method comprising:  
receiving a stream of words;  
15 employing a computer for generating a list of synonyms for at least one word in said stream of words;  
employing a computer for searching a corpus for synonym-containing phrases including at least one synonym in said list of synonyms together with at least part of said stream of words;  
20 employing a computer for evaluating the frequency of occurrence of each of said synonym-containing phrases; and  
proposing at least one selected synonym which forms part of a synonym-containing phrase having a relatively high frequency of occurrence in said corpus.
- 25 210. A computerized synonym generating method according to claim 209 and also comprising:  
employing a computer for searching said corpus for received phrases including said at least one word together with said at least part of said stream of words;  
employing a computer for comparing the frequency of occurrence of said  
30 received phrases in said corpus with the frequency of occurrence of said synonym-containing phrases; and

proposing at least one selected synonym which forms part of a synonym-containing phrase only if the frequency of occurrence of said synonym-containing phrase exceeds the frequency of occurrence of said received phrase.

5 211. A computerized synonym generating method according to either of claims 209 and 210 and wherein said at least one word comprises at least one of a noun, a verb, an object of a verb and a subject of a verb.

212. A computerized synonym generating system comprising:  
10 a computer operative to generate a list of synonyms for at least one word in a stream of words received from a user;  
computerized searching functionality operative to search a corpus for synonym-containing phrases including at least one synonym in said list of synonyms together with at least part of said stream of words;  
15 computerized frequency evaluation functionality operative to evaluate the frequency of occurrence of each of said synonym-containing phrases; and  
computerized synonym providing functionality operative to propose at least one selected synonym which forms part of a synonym-containing phrase having a relatively high frequency of occurrence in said corpus.

20 213. A computerized synonym generating system according to claim 212 and also comprising:  
computerized received phrases searching functionality operative to search said corpus for received phrases including said at least one word together with  
25 said at least part of said stream of words; and  
computerized occurrence frequency comparing functionality operative to compare the frequency of occurrence of said received phrases in said corpus with the frequency of occurrence of said synonym-containing phrases,  
said computerized synonym providing functionality being operative to  
30 propose at least one selected synonym which forms part of a synonym-containing phrase only if the frequency of occurrence of said synonym-containing phrase exceeds the frequency of occurrence of said received phrase.

214. A document searching system according to either of claims 212 and 213 and wherein said at least one word comprises at least one of a noun, a verb, an object of a verb and a subject of a verb.

5

215. A computerized question generation method comprising:  
identifying at least one theme word in a document;  
searching for previously asked questions containing said at least one  
theme word or having previously generated answers containing said at least one theme  
word; and  
presenting said previously asked questions.

10

216. A computerized question generation method according to claim 215 and also comprising, prior to said identifying, employing a computer to obtain said  
document from a user, and wherein said presenting comprises presenting said  
previously asked questions on said computer to said user.

15

217. A computerized question generation method according to either of claims 215 and 216 and wherein said identifying comprises carrying out statistical  
analysis of the frequency of occurrence of words in said document.

20

218. A computerized question generation method according to claim 217 and wherein said carrying out statistical analysis comprises:

for each word in said document, stemming said word to a corresponding  
root word;

25

generating a word occurrence frequency score for each different root  
word corresponding to a word in said document;

using said word occurrence frequency scores to calculate a document  
word occurrence frequency indicating score for said document; and

selecting a subset of words in said document including at least one word  
having a word occurrence frequency score which is greater than or equal to at least said  
document word occurrence frequency indicating score.

30

219. A computerized question generation method according to claim 218 and wherein said document word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

220. A computerized question generation method according to any of claims 215 to 219 and wherein said identifying at least one theme word comprises selecting, as said at least one theme word, at least one word having a word occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

221. A computerized question generation method according to any of claims 215 to 220 and wherein said carrying out statistical analysis also comprises, following said selecting a subset of words in said document:

calculating a subset word occurrence frequency indicating score; and selecting, as said at least one theme word, at least one of said subset of words having a word occurrence frequency score which is greater than or equal to said subset word occurrence frequency indicating score.

222. A computerized question generation method according to claim 221 and wherein said subset word occurrence frequency indicating score comprises at least one of an average of said word occurrence frequency scores of words in said subset of words and a median of said word occurrence frequency scores of words in said subset of words.

223. A computerized question generation system comprising:  
computerized theme word identifying functionality for identifying at least one theme word in a document;  
computerized previous answer searching functionality operative to search for previously asked questions containing said at least one theme word or having previously generated answers containing said at least one theme word; and

an output device for providing said previously asked questions.

224. A computerized question generation system according to claim 223 and wherein said computerized theme word identifying functionality is operative to carry  
5 out statistical analysis of the frequency of occurrence of words in said document.

225. A computerized question generation system according to claim 224 and wherein said computerized theme word identifying functionality comprises:  
computerized word stemming functionality, operative, for each word in  
10 said document, to stem said word to a corresponding root word;  
a word occurrence frequency score generator for generating a word occurrence frequency score for each different root word corresponding to a word in said document;  
computerized document word occurrence frequency indicating score  
15 calculating functionality operative to use said word occurrence frequency scores to calculate a document word occurrence frequency indicating score for said document;  
and  
computerized word selecting functionality operative to select a subset of words in said document including at least one word having a word occurrence  
20 frequency score which is greater than or equal to said document word occurrence frequency indicating score.

226. A computerized question generation system according to claim 225 and wherein said computerized document word occurrence frequency indicating score  
25 calculating functionality is operative to calculate said document word occurrence frequency indicating score by calculating at least one of an average of said word occurrence frequency scores and a median of said word occurrence frequency scores.

227. A computerized question generation system according to either of claims  
30 225 and 226 and wherein said computerized theme word identifying functionality is operative to select, as said at least one theme word, at least one word having a word

occurrence frequency score which is greater than or equal to twice said document word occurrence frequency indicating score.

228. A computerized question generation system according to any of claims  
5 225 to 227 and wherein said computerized theme word identifying functionality also comprises:

computerized subset word occurrence frequency indicating score  
calculating functionality, operative to calculate a subset word occurrence frequency  
indicating score; and

10 computerized theme word selection functionality operative to select, as  
said at least one theme word, at least one of said subset of words having a word  
occurrence frequency score which is greater than or equal to said subset word  
occurrence frequency indicating score.

15 229. A computerized question generation system according to claim 228 and  
wherein said computerized subset word occurrence frequency indicating score  
calculating functionality is operative to calculate said subset word occurrence frequency  
indicating score by calculating at least one of an average of said word occurrence  
frequency scores of words in said subset of words and a median of said word occurrence  
20 frequency scores of words in said subset of words.

230. A computerized editable report precursor generating method comprising:  
inputting at least one question into a computer;  
employing said computer to obtain at least one answer to said at least one  
25 question;  
storing said at least one answer to said at least one question;  
presenting said at least one question to said at least one answer in an  
editable form on said computer as an editable report precursor;  
archiving a multiplicity of said editable report precursors; and  
30 following said archiving, employing said multiplicity of editable report  
precursors to enhance said employing said computer.

231. A computerized editable report precursor generating method according to claim 230 and wherein said archiving includes archiving edited versions of said multiplicity of editable report precursors and wherein said edited versions are also employed to enhance said employing said computer.

5

232. A computerized editable report precursor generating method according to either of claims 230 and 231 and wherein said inputting comprises inputting said at least one question to said computer by at least one of:

typing said question;  
10 using a voice responsive input device;  
using a screen scraping functionality;  
using an email functionality;  
using an SMS functionality; and  
using an instant messaging functionality.

15

233. A computerized editable report precursor generating method according to any of claims 230 to 232 and wherein said employing said computer comprises:

employing computerized answer retrieving functionality to generate document search terms including at least one additional search term not present in said  
20 question, which said additional search term was acquired, prior to receipt by said computer of said question from said user, by said computerized answer retrieving functionality in response to said at least one question; and

operating computerized search engine functionality to access a set of documents in response to said question, based not only on at least one search term  
25 supplied by a user but also on said at least one additional search term provided by said at least one computerized answer retrieving functionality.

234. A computerized editable report precursor generating method comprising:  
inputting at least one desired report subject identifier into a computer;  
30 employing said computer to generate at least one question related to a desired subject identified by said at least one desired report subject identifier;

employing said computer to obtain at least one answer to said at least one question; and

presenting said at least one question to said at least one answer in an editable form on said computer, thereby providing an editable report precursor.

5

235. A computerized editable report precursor generating method according to claim 234 and also comprising:

archiving a multiplicity of said editable report precursors; and

10 following said archiving, employing said multiplicity of editable report precursors to enhance at least one of said employing said computer to generate at least one question and said employing said computer to obtain at least one answer to said at least one question.

236. A computerized editable report precursor generating method according to  
15 either of claims 234 and 235 and wherein said archiving includes archiving edited versions of said multiplicity of editable report precursors and wherein said edited versions are also employed to enhance at least one of said employing said computer to generate at least one question and said employing said computer to obtain at least one answer to said at least one question.

20

237. A computerized editable report precursor generating method according to any of claims 234 to 236 and wherein said inputting comprises inputting said at least desired report subject identifier to said computer by at least one of:

25 typing said desired report subject identifier;  
using a voice responsive input device;  
using a screen scraping functionality;  
using an email functionality;  
using an SMS functionality; and  
using an instant messaging functionality.

30

238. A computerized editable report precursor generating method according to any of claims 234 to 237 and wherein said employing said computer to generate said at



least one question comprises employing said desirable report subject identifier to search for previously asked questions containing at least part of said desirable report subject identifier or having previously generated answers containing at least part of said desirable report subject identifier.

5

239. A computerized editable report precursor generating method according to any of claims 234 to 238 and wherein said employing said computer comprises:

employing computerized answer retrieving functionality to generate document search terms including at least one additional search term not present in said question, which said additional search term was acquired, prior to receipt by said computer of said desired report subject identifier from said user, by said computerized answer retrieving functionality in response to at least one query;

operating computerized search engine functionality to access a set of documents in response to said question, based not only on said desired report subject identifier but also on said at least one additional search term provided by said at least one computerized answer retrieving functionality.